

⑩ 日本国特許庁(JP)

⑪ 特許出願公開

⑫ 公開特許公報(A)

昭63-316097

⑬ Int.Cl.<sup>4</sup>

G 10 L 3/00  
3/02

識別記号

3 0 1  
3 0 1

庁内整理番号

A-7627-5D  
7627-5D

⑭ 公開 昭和63年(1988)12月23日

審査請求 未請求 発明の数 1 (全4頁)

⑮ 発明の名称 連続音声認識装置

⑯ 特 願 昭62-152520

⑰ 出 願 昭62(1987)6月19日

⑱ 発 明 者 羽 金 廣 東京都港区芝5丁目33番1号 日本電気株式会社内

⑲ 出 願 人 日本電気株式会社 東京都港区芝5丁目33番1号

⑳ 代 理 人 弁理士 本庄 伸介

明 細 書

1. 発明の名称

連続音声認識装置

2. 特許請求の範囲

連続的に発声された音声認識する連続音声認識装置において、入力音声信号を互いに異なる増幅度でそれぞれ増幅する複数の増幅器と、前記入力音声信号の始端を検出する始端検出部と、前記入力音声信号の終端を検出する終端検出部と、前記始端から前記終端までの期間における複数の前記増幅器の出力をそれぞれ一時記憶する複数の記憶部と、前記始端から前記終端に到るまでの期間を複数のセグメントに分割する手段と、各前記セグメントごとに当該セグメントにおける最適な増幅度に対応する前記記憶部から音声信号を読み出して認識処理を行なう認識部とを有する連続音声認識装置。

3. 発明の詳細な説明

(産業上の利用分野)

本発明は、連続的に発声された音声認識する音声認識装置に関する。

(従来の技術)

従来、この種の連続音声認識装置では、登録(特定話者用の音声認識装置で行なう処理)や認識処理を行なう前に、マイクロホンからの音声信号を増幅するための増幅器の増幅度を発声者自身がボリューム等で設定したり、あるいは発声者のテスト発声のレベルに従って音声認識装置が自動的に増幅度を設定して(以後このレベルを設定する処理をレベル設定と呼ぶ)、その増幅度で増幅された音声信号に対して連続音声認識処理を行なっていた。

(発明が解決しようとする問題点)

上述した従来の連続音声認識装置では、レベル設定後に登録や認識が行なわれ、レベル設定で決定された増幅度は、その登録や認識処理中は変更されることなく一定の増幅度でマイクロホンから

の音声信号を増幅して処理している。このようにして増幅度を設定する従来の連続音声認識装置では、発声者の音量が変化した場合には最適な増幅度の音声信号が得られない場合があった。万一増幅度が最適となった場合でも連続的に発声された音声では音量が大きい認識対象区間に対して最適になるように増幅度が決められるから、連続的に発声された音声の中に音量が小さい認識対象区間が存在した場合、その区間では増幅度が小さい。(音量が大きい部分に比べてA/D変換後の量子化の精度が粗い)音声信号を処理して音声認識を行なわざるをえない欠点があった。具体的な例を第2図で説明する。

第2図は、発声者が「52」を発声したときのマイクからの音声信号パワー(音量)の変化を示す図である。一般的に「52」の発声の中で、「5」の発声区間は「2」の発声区間に対してパワーが大きい。「5」の発声区間に最適な増幅度をもった音声信号が万一得られた場合でもダイナミックレンジ(第2図のレベルD)に占める「2」の音

量のレベルの割合が小さいから、A/D変換後に十分な分解能が得られず、「2」の発声区間の量子化の精度が「5」のそれに比較して粗くなる欠点があった。

(問題点を解決するための手段)

前述の問題点を解決するために本発明が提供する手段は、連続的に発声された音声を認識する連続音声認識装置であって、入力音声信号を互いに異なる増幅度でそれぞれ増幅する複数の増幅器と、前記入力音声信号の始端を検出する始端検出部と、前記入力音声信号の終端を検出する終端検出部と、前記始端から前記終端までの期間における複数の前記増幅器の出力をそれぞれ一時記憶する複数の記憶部と、前記始端から前記終端に到るまでの期間を複数のセグメントに分割する手段と、各前記セグメントごとに当該セグメントにおける最適な増幅度に対応する前記記憶部から音声信号を読み出して認識処理を行なう認識部とを有してなる。

(実施例)

次に本発明の実施例について図面を参照して説明する。

第1図は本発明の連続音声認識装置の一実施例を示す構成図である。

本実施例において、増幅器 $A_1, A_2, \dots, A_n$ はそれぞれ異なる増幅度をもっており、マイクロホンMCから入力した音声信号をそれぞれ増幅する。始端検出部SD、終端検出部EDは音声信号の始めと終わりを検出しそれぞれ始端検出信号 $S_1$ 、終端検出信号 $S_2$ を出力する。A/D変換器 $C_1, C_2, \dots, C_n$ は増幅器 $A_1, A_2, \dots, A_n$ の出力をそれぞれデジタルな音声信号に変換する。記憶部 $B_1, B_2, \dots, B_n$ はそれぞれ始端検出信号 $S_1$ を入力してから終端検出信号 $S_2$ を入力するまでA/D変換器 $C_1, C_2, \dots, C_n$ からの音声信号を記憶する。連続音声認識部(以後認識部と呼ぶ)RCは、終端検出信号 $S_2$ を受信した時点で、連続音声認識処理を開始する。

認識部RCの処理を第3図の例を使って説明する。第3図は連続的に発声されたある音声の音量

の変化を示した図である。認識部RCは連続的に発声された音声に仮区分点P、Qを設定して3つのセグメントA、B、Cに分割する。認識部RCはセグメントAの始点 $a_1$ をセグメント始点信号 $S_1$ 、終点 $a_2$ をセグメント終点信号 $S_2$ として増幅度選択部SEへ送る。増幅度選択部SEは、始点 $a_1$ と終点 $a_2$ で示されるセグメントAの区間に対して最適な増幅度で記憶されている音声信号を記憶部 $B_1 \sim B_n$ の中から選び、セグメント音声信号Vとして認識部RCへ送る。認識部RCは送られてきたセグメントAの音声信号から特徴を抽出して、その特徴からセグメントAの認識結果Aを得る。具体的には、バンドパスフィルターの出力を特徴として、標準ボタンと未知ボタン(セグメントAの特徴から得られたボタン)の一致の度合(類似度)を調べてその類似度が最も大きい標準ボタンのカテゴリ(属性)を認識結果とするボタンマッチング法で音声認識が実現できる。第3図のセグメントB、Cについても同様の処理を行なって認識結果B、Cを得る。

ここで得られた認識結果A, B, Cは、連続的に発声された音声を仮の区分点P, Qで3分割した場合の結果であり最終的な認識結果を得るための候補となる。認識部RCは、連続音声のセグメント(分割)数とそれに伴う区分点の位置を変数としてこのすべての変数の組合せに対して第3図の例で説明した処理をくり返し行ない、各々の組合せ毎に認識結果の候補を得て、その候補の中から最も類似度の大きい候補を最終的な連続音声の認識結果Tとして出力する。

具体的には、認識部RCの連続認識の処理は前述したパターンマッチング法とDP法(DYNAMIC PROGRAMING法)を用いて実現できる。

以上、認識部RCの一実施例としてパターンマッチング法、DP法をあげたが、他のいかなる認識方式についても本発明が適用できることは明らかである。また本実施例では増幅度の段階を5レベルとして説明したが、本発明は5レベルに限定されるものではない。

(発明の効果)

器、 $C_1, C_2, \dots, C_n$ はA/D変換器、 $M_1, M_2, \dots, M_n$ は記憶部、SEは増幅度選択部、RCは連続音声認識部、SDは始端検出部、EDは終端検出部、 $S_1$ は始端検出信号、 $S_2$ は終端検出信号、Vはセグメント音声信号、 $S_3$ はセグメント始点信号、 $S_4$ はセグメント終点信号、Tは連続音声認識結果である。

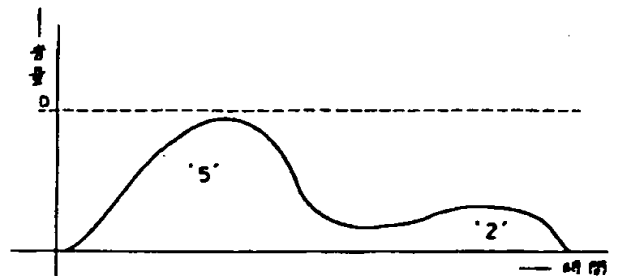
代理人 弁理士 本庄伸介

以上説明したように本発明では、連続音声を複数のセグメントに分割して各々のセグメントの認識処理を行なう場合、各セグメント毎にそのセグメントの区間によって最適な増幅度をもった音声信号を記憶部より選択するので、発声レベルが低い連続音声の部分に対しても最適な増幅度で増幅された音声信号の処理が可能である。そこで、本発明の連続音声認識装置によれば、従来発声レベルが低いために特徴量が十分に抽出されないことが原因となって発生したエラーやリジェクトを削減できる。

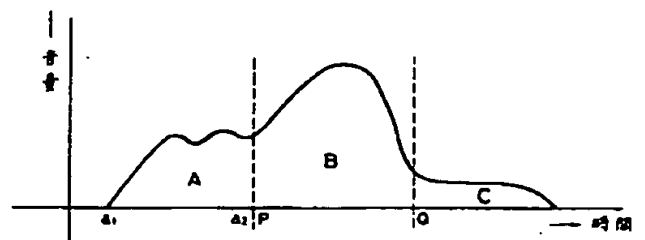
#### 4. 図面の簡単な説明

第1図は本発明の連続音声認識装置の一実施例を示す構成図、第2図は発声者が「52」を発声したときのマイクからの音声信号の音量の変化を示す図、第3図は連続的にある音声を発声したときのマイクからの音声信号の音量の変化を示す図である。

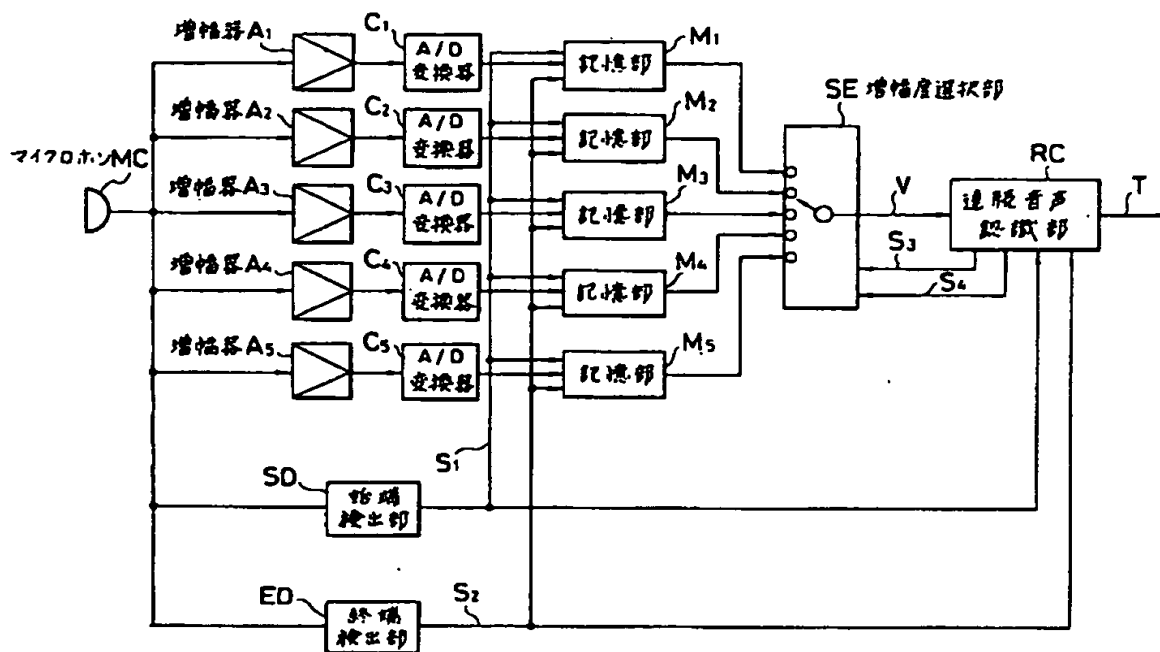
MCはマイクロホン、 $A_1, A_2, \dots, A_n$ は増幅



第2図



第3図



第 1 図

Partial Translation of JP63-316097

[Problem to be Solved by the Invention]

In the aforementioned conventional continuous voice  
5 recognizer, registration and recognition are performed after  
level setting and the degree of amplification decided in the  
level setting is not changed during the registration and  
recognition processing but the voice signal from the microphone  
is amplified and processed at the constant degree of  
10 amplification. In the conventional continuous voice  
recognizer setting the degree of amplification in the  
aforementioned manner, no voice signal of the optimum degree  
of amplification may be obtained when the volume of the speaker  
changes. The degree of amplification is decided to be optimum  
15 for a recognized section having large volume in a continuously  
generated voice even if the degree of amplification is  
optimized, and hence, when a recognized section having small  
volume is present in the continuously generated voice, the  
degree of amplification is small in this section. The voice  
20 signal (having rougher precision of quantization after A/D  
conversion as compared with a part having large volume) must  
be disadvantageously processed for performing voice  
recognition. A specific example is now described with  
reference to Fig. 2.

25 Fig. 2 is a diagram showing change of voice signal power

(volume) received from a microphone when a speaker vocalizes "52". In the vocalization of "52", the vocalization section "5" generally has larger power than the vocalization section "2". The rate of the level of the volume of "2" occupying the dynamic range (level D in Fig. 2) is small even if a voice signal having the optimum degree of amplification for the vocalization section "5" is obtained, no sufficient resolution is obtained after A/D conversion but precision of quantization of the vocalization section "2" is disadvantageously roughened as compared with that of "5".

(Means for Solving the Problem)

In order to solve the aforementioned problem, the present invention provides a continuous voice recognizer recognizing continuously generated voice, which has a plurality of amplifiers amplifying an input voice signal at different degrees of amplification respectively, a beginning end detector detecting the beginning end of the said input voice signal, a terminating end detector detecting the terminating end of the said input voice signal, a plurality of storages temporarily storing outputs from the plurality of said amplifiers in the period between the said beginning end and the said terminating end respectively, means dividing the period between the said beginning end and the said terminating end into a plurality of segments, and a recognizer reading the voice signal from the said storage corresponding to the optimum

degree of amplification in each of the said segments and performing recognition processing every said segment.

## ⑫ 公開特許公報(A)

昭63-316097

⑪ Int. Cl.<sup>4</sup>G 10 L 3/00  
3/02

識別記号

3 0 1  
3 0 1

庁内整理番号

A-7627-5D  
7627-5D

⑬ 公開 昭和63年(1988)12月23日

審査請求 未請求 発明の数 1 (全4頁)

⑭ 発明の名称 連続音声認識装置

⑮ 特 願 昭62-152520

⑯ 出 願 昭62(1987)6月19日

⑰ 発 明 者 羽 金 廣 東京都港区芝5丁目33番1号 日本電気株式会社内

⑱ 出 願 人 日本電気株式会社 東京都港区芝5丁目33番1号

⑲ 代 理 人 弁理士 本庄 伸介

## 明 細 書

## 1. 発明の名称

連続音声認識装置

## 2. 特許請求の範囲

連続的に発声された音声認識する連続音声認識装置において、入力音声信号を互いに異なる増幅度でそれぞれ増幅する複数の増幅器と、前記入力音声信号の始端を検出する始端検出部と、前記入力音声信号の終端を検出する終端検出部と、前記始端から前記終端までの期間における複数の前記増幅器の出力をそれぞれ一時記憶する複数の記憶部と、前記始端から前記終端に到るまでの期間を複数のセグメントに分割する手段と、各前記セグメントごとに当該セグメントにおける最適な増幅度に対応する前記記憶部から音声信号を読み出して認識処理を行なう認識部とを有する連続音声認識装置。

## 3. 発明の詳細な説明

(産業上の利用分野)

本発明は、連続的に発声された音声認識する音声認識装置に関する。

(従来の技術)

従来、この種の連続音声認識装置では、登録(特定話者用の音声認識装置で行なう処理)や認識処理を行なう前に、マイクロホンからの音声信号を増幅するための増幅器の増幅度を発声者自身がボリューム等で設定したり、あるいは発声者のテスト発声のレベルに従って音声認識装置が自動的に増幅度を設定して(以後このレベルを設定する処理をレベル設定と呼ぶ)、その増幅度で増幅された音声信号に対して連続音声認識処理を行っていた。

(発明が解決しようとする問題点)

上述した従来の連続音声認識装置では、レベル設定後に登録や認識が行なわれ、レベル設定で決定された増幅度は、その登録や認識処理中は変更されることなく一定の増幅度でマイクロホンから



の音声信号を増幅して処理している。このようにして増幅度を設定する従来の連続音声認識装置では、発声者の音量が変化した場合には最適な増幅度の音声信号が得られない場合があった。万一増幅度が最適となった場合でも連続的に発声された音声では音量が大きい認識対象区間に対して最適になるように増幅度が決められるから、連続的に発声された音声の中に音量が小さい認識対象区間が存在した場合、その区間では増幅度が小さい。(音量が大きい部分に比べてA/D変換後の量子化の精度が粗い)音声信号を処理して音声認識を行なわざるをえない欠点があった。具体的な例を第2図で説明する。

第2図は、発声者が「52」を発声したときのマイクからの音声信号パワー(音量)の変化を示す図である。一般的に「52」の発声の中で、「5」の発声区間は「2」の発声区間に比べてパワーが大きい。「5」の発声区間に最適な増幅度をもった音声信号が万一得られた場合でもダイナミックレンジ(第2図のレベルD)に占める「2」の音

量のレベルの割合が小さいから、A/D変換後に十分な分解能が得られず、「2」の発声区間の量子化の精度が「5」のそれに比較して粗くなる欠点があった。

(問題点を解決するための手段)

前述の問題点を解決するために本発明が提供する手段は、連続的に発声された音声認識する連続音声認識装置であって、入力音声信号を互いに異なる増幅度でそれぞれ増幅する複数の増幅器と、前記入力音声信号の始端を検出する始端検出部と、前記入力音声信号の終端を検出する終端検出部と、前記始端から前記終端までの期間における複数の前記増幅器の出力をそれぞれ一時記憶する複数の記憶部と、前記始端から前記終端に至るまでの期間を複数のセグメントに分割する手段と、各前記セグメントごとに当該セグメントにおける最適な増幅度に対応する前記記憶部から音声信号を読み出して認識処理を行なう認識部とを有してなる。

(実施例)

次に本発明の実施例について図面を参照して説明する。

第1図は本発明の連続音声認識装置の一実施例を示す構成図である。

本実施例において、増幅器 $A_1, A_2, \dots, A_n$ はそれぞれ異なる増幅度をもっており、マイクロホンMCから入力した音声信号をそれぞれ増幅する。始端検出部SD、終端検出部EDは音声信号の始めと終わりを検出しそれぞれ始端検出信号 $S_1$ 、終端検出信号 $S_2$ を出力する。A/D変換器 $C_1, C_2, \dots, C_n$ は増幅器 $A_1, A_2, \dots, A_n$ の出力をそれぞれデジタルな音声信号に変換する。記憶部 $B_1, B_2, \dots, B_n$ はそれぞれ始端検出信号 $S_1$ を入力してから終端検出信号 $S_2$ を入力するまでA/D変換器 $C_1, C_2, \dots, C_n$ からの音声信号を記憶する。連続音声認識部(以後認識部と呼ぶ)RCは、終端検出信号 $S_2$ を受信した時点で、連続音声認識処理を開始する。

認識部RCの処理を第3図の例を使って説明する。第3図は連続的に発声されたある音声の音量

の変化を示した図である。認識部RCは連続的に発声された音声に仮区分点P、Qを設定して3つのセグメントA、B、Cに分割する。認識部RCはセグメントAの始点 $a_1$ をセグメント始点信号 $S_1$ 、終点 $a_2$ をセグメント終点信号 $S_2$ として増幅度選択部SEへ送る。増幅度選択部SEは、始点 $a_1$ と終点 $a_2$ で示されるセグメントAの区間に対して最適な増幅度で記憶されている音声信号を記憶部 $B_1 \sim B_n$ の中から選び、セグメント音声信号Vとして認識部RCへ送る。認識部RCは送られてきたセグメントAの音声信号から特徴を抽出して、その特徴からセグメントAの認識結果Aを得る。具体的には、バンドパスフィルターの出力を特徴として、標準ボタンと未知ボタン(セグメントAの特徴から得られたボタン)の一致の度合(類似度)を調べてその類似度が最も大きい標準ボタンのカテゴリ(属性)を認識結果とするボタンマッチング法で音声認識が実現できる。第3図のセグメントB、Cについても同様の処理を行なって認識結果B、Cを得る。

ここで得られた認識結果A, B, Cは、連続的に発声された音声を仮の区分点P, Qで3分割した場合の結果であり最終的な認識結果を得るための候補となる。認識部RCは、連続音声のセグメント(分割)数とそれに伴う区分点の位置を変数としてこのすべての変数の組合せに対して第3図の例で説明した処理をくり返し行ない、各々の組合せ毎に認識結果の候補を得て、その候補の中から最も類似度の大きい候補を最終的な連続音声の認識結果Tとして出力する。

具体的には、認識部RCの連続認識の処理は前述したパターンマッチング法とDP法(DYNAMIC PROGRAMMING法)を用いて実現できる。

以上、認識部RCの一実施例としてパターンマッチング法、DP法をあげたが、他のいかなる認識方式についても本発明が適用できることは明らかである。また本実施例では増幅度の段階を5レベルとして説明したが、本発明は5レベルに限定されるものではない。

(発明の効果)

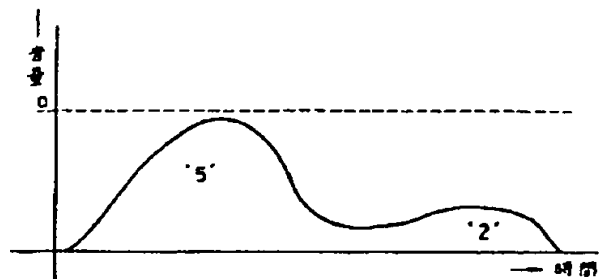
器、 $C_1, C_2, \dots, C_n$ はA/D変換器、 $M_1, M_2, \dots, M_n$ は記憶部、SEは増幅度選択部、RCは連続音声認識部、SDは始端検出部、EDは終端検出部、 $S_1$ は始端検出信号、 $S_2$ は終端検出信号、Vはセグメント音声信号、 $S_3$ はセグメント始点信号、 $S_4$ はセグメント終点信号、Tは連続音声認識結果である。

以上説明したように本発明では、連続音声を複数のセグメントに分割して各々のセグメントの認識処理を行なう場合、各セグメント毎にそのセグメントの区間にとって最適な増幅度をもった音声信号を記憶部より選択するので、発声レベルが低い連続音声の部分に対しても最適な増幅度で増幅された音声信号の処理が可能である。そこで、本発明の連続音声認識装置によれば、従来発声レベルが低いために特徴量が十分に抽出されないことが原因となって発生したエラーやリジェクトを削減できる。

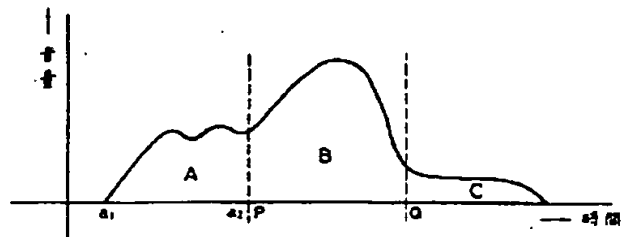
#### 4. 図面の簡単な説明

第1図は本発明の連続音声認識装置の一実施例を示す構成図、第2図は発声者が「52」を発声したときのマイクからの音声信号の音量の変化を示す図、第3図は連続的にある音声を発声したときのマイクからの音声信号の音量の変化を示す図である。

MCはマイクロホン、 $A_1, A_2, \dots, A_n$ は増幅

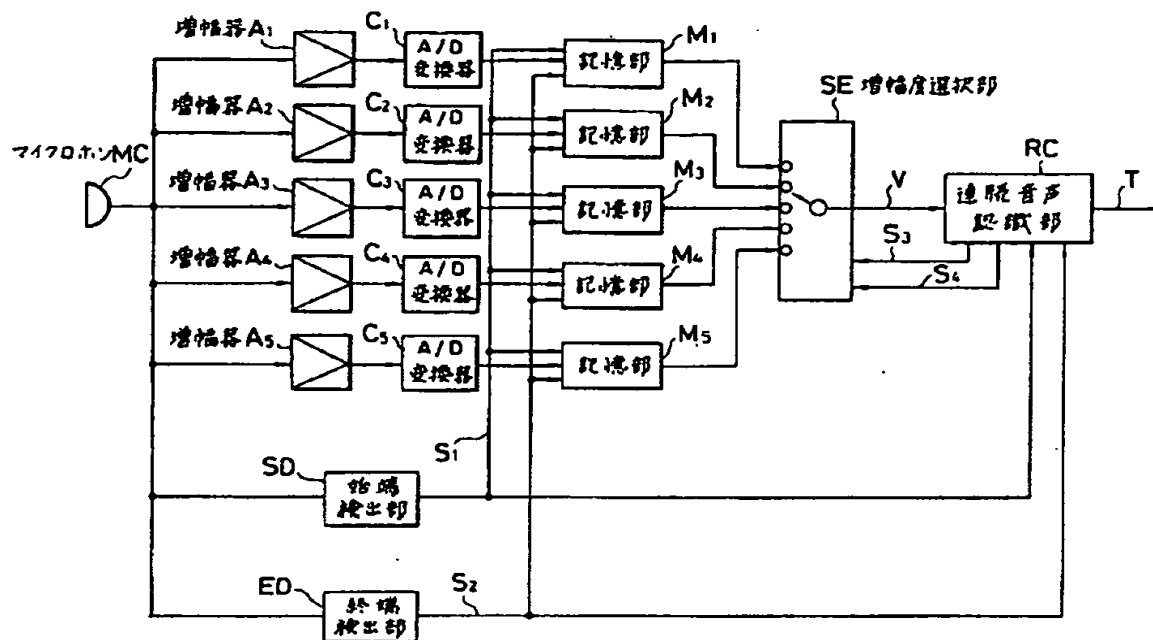


第2図



第3図

代理人 弁理士 本庄伸介



第 1 図